DOCUMENT RESUME

ED 295 994 TM 011 849

AUTHOR Elliot, Scott M.; And Others
TITLE Updating Teacher Certification Tests: Procedures and Outcomes.
INSTITUTION National Evaluation Systems, Inc., Amherst, Mass.
PUB DATE Apr 88
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 6-8, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Methods; *Item Banks; *Licensing Examinations (Professions); State Licensing Boards; State Programs; *Teacher Certification; Teacher Evaluation; Test Bias; *Test Construction; Testing Programs; Test Reviews; Test Validity
IDENTIFIERS *Teacher Competency Testing; Texas Education Agency

ABSTRACT

The process used by the Texas Education Agency to update public school teacher certification tests in a statewide program in Texas is described. The testing program implemented is called the Examination for the Certification of Educators in Texas. After a policy review, advisory committees for 63 test areas prepared content frameworks, created objectives, and reviewed job analysis surveys prepared by classroom teachers and educators. A bank of test items was written for each test area and field tested at colleges and universities throughout the state. Minority educators reviewed the tests for bias. Recommended passing scores were obtained from teachers. The procedures described offer a useful model for keeping teacher certification tests up to date. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

UPDATING TEACHER CERTIFICATION TESTS: PROCEDURES AND OUTCOMES

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

SCOTT M. ELLIOT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

UPDATING TEACHER CERTIFICATION TESTS: PROCEDURES AND OUTCOMES

Scott M. Elliot
Paula M. Nassif
Sharon L. Downs

National Evaluation Systems, Inc.
30 Gatehouse Road
P. O. Box 226
Amherst, Massachusetts 01004

Paper presented at the annual conference of the National Council on Measurement in Education, New Orleans, 1988.

UPDATING TEACHER CERTIFICATION TESTS: PROCEDURES AND OUTCOMES

The use of tests as one criterion for certifying prospective teachers has become a common practice; close to forty states now make use of such measures in their certification process. This experience has given the field a significant body of information to guide the initial development of these tests. Yet, there is far less available to guide test developers in updating and revising these tests.

The field of education ranging from teacher educator training practices to the knowledge base in some teaching fields is changing constantly. In order to meet the demands of a dynamic teacher certification testing program, test developers, and state departments of education need a technically sound and practical process for updating teacher certification tests.

This paper describes the process used to update teacher certification tests in a large-scale, statewide teacher certification testing program. The methodology used to update the tests, the outcomes of the updating effort and the implications of the results are described.

PROGRAM OVERVIEW

Program Background

In 1981, the Texas legislature passed Senate Bill 50 relating to the certification of public school teachers and other educational personnel. One component of this legislation provided for the development and administration of examinations in areas in which teacher certification is granted. The tests are required for individuals seeking initial certification, and for those educators seeking additional certificate or endorsement areas. The testing program that has been implemented in response to this legislation is the Examination for the Certification of Educators in Texas (ExCET).

The Texas Education Agency (TEA) with the assistance of its selected contractor, developed sixty-three criterion-referenced tests covering sixty certificate areas for the ExCET Program. Thirty-four of these tests have been administered since May 1986; the remaining twenty-nine have been administered since June 1987.

Initial Test Development Process

The initial test development process incorporated a series of steps described below.

1. Policy review. Relevant Texas laws, the Texas Essential Elements, teacher education program standards, adopted textbooks, and other instructional policies and documents were reviewed in preparation for test development.

2. Advisory committees. In each of the 63 test areas advisory committees, composed of Texas public school teachers and teacher educators, were formed to review materials for content and equity issues, and to assist in setting passing scores for the tests. Three separate groups were assembled: 1) test development advisory committees, 2) bias reviewers and 3) standard setting participants.

3. Content frameworks. A content framework (or content outline) was prepared for each content area.

4. Objectives. Objectives were created based on the content framework for each content area.

5. Framework and objective review. Each advisory committee reviewed and revised the content framework and objectives for its test area.

6. Objective correlation study. Each approved test objective was correlated with one or more Texas Essential Elements, teacher education program standards, or adopted textbooks.

7. Job analysis surveys. Thousands of Texas classroom teachers and teacher educators completed job analysis surveys in which they rated the importance of each proposed objective to the classroom responsibilities of teachers. In addition, teacher education program students were asked in a separate survey to indicate whether they had an opportunity to learn the content described by the objectives.

8. Objective selection. The results of the job analysis surveys in each content area were reviewed by the advisory committees; using these results, the committees selected the objectives to be tested.

9. Item specifications. Test item specifications describing the specific content to be measured by each test objective were written. The advisory committees reviewed and revised the item specifications.

10. Test items. A bank of test items was written for each test area based on the approved objectives and the item specifications. The advisory committees reviewed and revised the test items that were prepared.

11. Field test. All test items were field tested at colleges and universities throughout the state. The advisory committees reviewed the field test results and used them to refine the test items.

12. Bias review. All test items were reviewed for potential bias by minority educators in Texas. Field test results and item statistics were used in flagging items for potential bias. Bias review also occurred at each developmental stage of the project.

13. Content validation. A second group of Texas classroom teachers and teacher educators was convened to validate further the content of the test items.

14. Recommended passing scores. Texas classroom teachers and teacher educators reviewed the tests and recommended passing scores.

15. Set passing scores. Final passing scores were set by the State Board of Education.

As part of the ongoing activities of the ExCET Program, certification tests are reviewed and updated on a periodic basis, as necessary. The test updating process is designed to keep the tests up to date by verifying that the tests continue to reflect job requirements, state policy, and other features of the field (e.g., terminology, knowledge base).

TEST UPDATING PROCEDURES

In the summer and fall of 1987, sixteen tests were updated for the ExCET Program. Three primary activities were carried out in the updating process:

- Bias Review
- Item Development and Review
- Minimum Passing Score Review

These activities are described following a description of the committees involved in the test updating activities.

Bias Review and Test Update Advisory Committees

To assist in the test updating process, two separate committees were formed by the Texas Education Agency. First, a panel of approximately thirty representatives from various cultural and ethnic groups represented in the state were assembled; some had performed the original bias review during initial development and some were newly added to the review. Their charge was to review materials for potential bias. Second, a test updating advisory committee composed of ten to twenty members, consisting of a core from the original committee plus some newly added, was convened to review test materials in each field. The bias reviewers met for one day, while the test update

advisory committees met for one or two days depending on the length of the test in the area they were reviewing. The committees received orientation and training including a complete overview of the purpose and history of the program, the purpose of the test updating and a description of the test updating procedures.

Bias Review

Preventing bias must be a central focus of any teacher certification test development effort. Review of items for potential bias by judges as well as statistical item bias analyses may be considered according to the 1985 standards for educational and psychological testing of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. The importance of reviewing test items for potential bias is echoed by many researchers and test developers in the field (c.f. Berk, 1982). While few would deny the importance of taking reasonable measures to prevent bias in teacher certification tests, there is less consensus on what techniques should be applied to prevent bias. Berk (1986) suggests that bias review should not be a static event, but should be an ongoing process involving at least three components: 1) judgmental review or "content bias", 2) statistical item bias, and 3) a posteriori analysis for bias. Berk (1986) notes "that this scheme involves: first, judgments; second, statistics; and third, judgments."

The bias review component of the updating process for the sixteen fields examined for the ExCET program applied both judgmental .eview by educators and the application of statistical bias analyses. As the first step in the bias review for the test updating, all items in the current bank that had appeared on a test to date were statistically analyzed for bias using the operational test administration data. The delta plot method described by Angoff (1982) with the adjustments suggested by Shepard, Camilli, and Williams (1985) was applied to examine items statistically for bias. The modifications suggested by Shepard, et. al. (1985) calls for the computation of a regression between the simple delta value yielded from the Angoff delta plot and the item-total point biserial correlation, and using the residuals to detect potential item bias. Bias analysis was conducted by test form.

The design for the program calls for test forms to change periodically; if an item was flagged based on the statistical bias analysis procedure on any test form administered, it was highlighted for further review. Approximately 10% of the items across the sixteen fields were flagged for review based on these procedures.

The second major task was the judgmental review of test items for potential bias by bias review panel members. The bias reviewers were asked to review those items within the current banks that had been flagged for potential bias based on the residualized delta bias analysis.

Bias reviewers were provided with the test objectives for which test items were written, a complete set of item statistics for all items that had appeared on a test form to date (including p-values, point biserial correlations, the distribution of responses, the p-value for each subgroup of the population under consideration, and flags indicating which items had been flagged based on the statistical bias analysis), the bank of actual test items, and a training manual discussing item bias and the criteria for reviewing items for bias including language bias, content bias, stereotyping, and representation. Bias reviewers were asked to review each item in the current bank and to indicate for each flagged item whether or not he or she felt the item was biased based on the criteria provided. If he or she felt there was potential bias the reviewer was asked to indicate the reason for making that judgment and recommendations for correcting the item. Following the review of flagged items within the existing item banks, reviewers were also asked to review all newly developed items proposed for inclusion on future test forms for potential bias. The results of the bias review, along with the recommendations of the bias reviewers for revising items were presented to the test update advisory committees in each test area for their consideration in reviewing the items.

Item Review

Following the Bias Review, the test updating advisory committees were asked to review both the items in the current item bank and a set of newly developed items for use on future forms of the test.

Review of items in current bank. All test items in the current bank were reviewed by the test update advisory committees. Committee members were provided with an orientation and training and then asked to review items for:

- Topicality
- Item Performance
- Examinee and Other Reviewer Comments
- Bias Review Comments

To assist committee members in reviewing item performance, each committee member was provided with:

- cumulative item statistics for all test administrations to date; these data included p-values, point-biserial correlations, response distributions, and p-value information for relevant subgroups of the population.

- the objectives for which test items were written, and

- a booklet containing the bank of current items.

Any item with a p-value less than or equal to .40, or with an item to test point biserial correlation less than or equal to .10 was flagged for committee review; approximately 20% of the items across the sixteen fields were flagged for further review based on these criteria. Committee members reviewed all items in the bank and revised, if necessary, statistically flagged items, those with potential topicality problems, items questioned by examinees, and those highlighted by the bias review panel.

Review of new test items. For each test area, between fifty and one hundred new items were written to match the test objectives. This allowed the introduction of new content and the augmentation of the current item bank to allow the creation of new test forms. The test update advisory committee reviewed each new item for objective match, accuracy, freedom from bias, and job-relatedness and made revisions as necessary. The committee also reviewed comments on the newly developed items indicated by the bias review panel.

Minimum Passing Score Review

Since the mid-1950s, numerous procedures for establishing the minimum passing scores on tests have been proposed. Comprehensive review of available standard setting methods is provided by Berk (1986) and a description of their application to teacher certification tests is provided by Nassif (1986).

Initial Standard Setting Process. As part of the initial test development process for the ExCET program, minimum passing scores were established using a standard setting process where judgments were systematically collected from professional educators in the state, the judgments were summarized, and then reported to the state board of education for their determination of the passing score for each test. The initial standard setting procedure combined features of both the Angoff (1971) and Jaeger (1982) models. Judges provided estimates of the percentage of minimally competent persons that would answer each question correctly. And, consistent with the standard setting procedure described by Jaeger (1982), judges were given a second opportunity to reconsider their ratings based on a summary of the first round of ratings provided. During the first round judges reviewed the individual test items, reviewed the field test data on individual items, and finally rated the individual test items. During the second round, judges were provided with their original ratings and summary data for all judges in the field including the range of ratings for each item and the median rating for each item. Then, judges were asked to review their initial rating of individual test items and retain or revise their original rating. The results from the second round of ratings were compiled and information was provided to the Texas State Board of Education for their final determination of the test standards.

Test Updating Review of Standards. The procedures used to review the minimum passing scores for the ExCET program reflect an extension of the initial standard setting procedures. Judges on the test updating review committee were provided with an orientation and training session. The orientation and training session comprised a description of project background, an explanation of the standard setting process, and specific instructions on how to complete the standard setting procedures. Each committee member received:

- a training manual,
- a copy of the current test form,
- a summary statistics report indicating the current minimum passing score, along with the number of individuals passing the test overall and within each relevant subgroup of the population. The summary statistics report also included the distribution of scores on either side of the current minimum passing score and the percentage of individuals meeting or exceeding each score point both within the total group and each relevant subgroup of the population,
- the set of objectives for which test items were written, and
- a minimum passing score recommendation form on which to record the recommended passing score.

The review of the minimum passing scores followed the following steps as listed and described below:

- Orientation and Training,
- Discussion of acceptable standards for proficiency in the field,
- Review of the test form and summary statistics,
- Committee discussion and consensus on recommended minimum passing scores, and
- Final determination of minimum passing scores by the Board of Education.

Following the orientation, each committee was first asked to discuss the requirements for an entry-level educator in their field performing at an acceptable level in order to clarify acceptable standards for proficiency in their field. Committee members were then asked to review the test form to familiarize themselves with the test content and difficulty level of the test. They were also asked to review the summary statistics report to gain further information about the difficulty level of the test and how examinees had performed on the test to date. After all committee members had had an opportunity to familiarize themselves with the test form and current performance on

the test, the committee was asked to discuss their judgment of what constitutes an appropriate passing score for an entry-level educator meeting the standards for acceptable performance in the field. Specifically, they were asked to address the question: "what is the minimum percentage of items that an entry-level educator in the Texas public schools should be required to answer correctly in order to pass the ExCET test?" The committee was informed that they could elect to retain the current minimum passing score for the test or another minimum passing score. The committee was instructed to come to consensus on a recommended minimum passing score based on these considerations. The committee chairperson recorded the committee's final recommended passing score on the minimum passing score recommendation form provided.

The recommendations were provided to the Texas Education Agency for consideration and the Texas Education Agency made final recommendations on minimum passing scores to the board. In November 1987, the Texas State Board of Education determined the final minimum passing scores to be applied on the updated test forms.For nine of the fields the minimum passing score was not changed. Six of the fields cutscores increased and one of the fields decreased from the original cutscore levels.

This method provides a useful procedure for reviewing the initial standards established for tests. First, by having judges review an actual test form administered, a direct referent for the minimum passing score is established. Second, through guided discussion, judges are able to develop a clear understanding of the "standards for acceptable performance in the field." Finally, providing examinee performance data enabled judges to consider the impact of their decisions on the examinee population.

Test Form Construction

Following the conference the test item banks were revised to reflect the recommendations of the test update advisory committees. New test forms were produced in accordance with blueprinting guidelines for the ExCET Program. New test forms were created using both items from the previous test form and items newly revised or developed through the test updating. Sufficient overlap between the new and previous test form was maintained to permit equating. The new test form was equated to the form reviewed by the committee using the Tucker linear equating model.

Conclusion

Teacher certification tests are one important tool for determining whether an individual should or should not be permitted to enter the profession. Their role in this decision-making process is a critical one and as such reasonable measures should be taken to ensure that the decisions made on the basis of the tests are valid ones. Validity is not a static event; the tests should be reviewed periodically, as necessary, to maintain validity. The procedures described in this paper offer test developers a useful model for keeping teacher certification tests up-to-date, valid instruments for decision-making.

0296U

REFERENCES

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. C. Thorndike (Ed.) Educational Measurement (2nd ed.) (pp 508-600). Washington, D.C.: American Council of Education

Berk, R. A. (1982). Handbook of Methods of Detecting Test Bias. Baltimore, Md.: Johns Hopkins University Press.

Berk, R. A. (1986). Detecting bias in teacher certification tests, in Trends in Teacher Certification Testing. National Evaluation Systems, Inc.

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion referenced tests. Review of Educational Research, 56 137-172.

Jaeger, R. M. (1982). An alternative structural judgment process for establishing standards on competency tests: Theory and application. Educational Evaluation and Policy Analysis, 4, 461-475.

Nassif, P. M. (1986). Teacher certification testing technical challenger: Part I. in Testing for Teacher Certification, W. P. Gorth and M. L, Chernoff (eds.), Hilladale , N.J.: Lawrence Erlbaum Associates.

Shepard, L. A.; Camilli, G.; and Williams, P. M. (1985). Validity of approximation techniques for detecting item bias. Journal of Educational Measurement, 22, 77-105.